AD-A081 921 LOGOS DEVELOPMENT CORP MIDDLETOWN N Y F/G 9/2
DEVELOPMENT OF LOGOSCAN II.(U)
DEC 79 C E BYRNE, J C BYRNE, D J BYRNE F30602-78-C-0361
UNCLASSIFIED RADC -TR-79-315 NL

1 OF 1
AD
A081921

END
DATE
FILMED
4-80
DTIC

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14 **REPORT DOCUMENTATION PAGE**		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 RADC TR-79-315	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 DEVELOPMENT OF LOGOSCAN II.		5. ~~TYPE OF REPORT & PERIOD COVERED~~ 9 Final Technical Report, Oct~~ober~~ 78 – Aug~~ust~~ 1979,
		6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) 10 Charles E. Byrne John C. Byrne Daniel J. Byrne		8. CONTRACT OR GRANT NUMBER(s) 15 F30602-78-C-0361
9. ~~PERFORMING~~ ORGANIZATION NAME AND ADDRESS Logos Development Corporation 2 Low Avenue Middletown NY 10940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 31025 21830416 12 32
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRDT) Griffiss AFB NY 13441		12. ~~REPORT DATE~~ 11 Dec~~ember~~ 1979
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same 16 2183 17 04		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: John A. Guillen, Lt, USAF (IRDT)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Optical Character Reader
Cyrillic
Font Acquisition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this ~~effort~~ Rpt was to test and evaluate the Logoscan II Optical Character Reader System to assess its potential for conversion of free formatted multifont typeset Russian text to a computer processable format compatible with the Russian-English machine translation system at Foreign Technology Division (FTD). Loganscan II scanned 27 pages of Russian text supplied by FTD at approximately 30 characters per second with an error rate of 1.0 - 2.0%. This report discusses the methodology used, technical problems, results, and manpower cost for a production system.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

EVALUATION

The purpose of this effort was to test and evaluate the LOGOSCAN II Optical Character Reader System to assess its potential for conversion of free formatted multifont typeset Russian text to a computer processable format compatible with the Russian-English machine translation system at Foreign Technology Division (FTD).

LOGOSCAN II scanned 27 pages of Russian text supplied by FTD. The system scanned the text at approximately 30 characters per second with an error rate of 1.0 - 2.0%. The resultant study demonstrated that the scanning of Cyrillic text is feasible, and that the difficulties encountered in scanning are due to difficulties inherent in Cyrillic text.

LOGOSCAN II is currently not a production system. Further work in the areas of scanning speed, recognition accuracy and post editor programs to correct errors is needed for such a system.

John A. Guillen

Project Engineer

A. INTRODUCTION

The optical scanning of typeset Cyrillic text is perhaps the ultimate test of an OCR system's recognition algorithm. The problem divides logically into four parts:

1. recognizing the character boundaries;
2. differentiating the character from all others in a reference alphabet;
3. tracking a skewed line, and recognizing line boundaries;
4. selecting the correct reference alphabet from a set of possible reference alphabets.

The Logoscan II System solves all four of these problems.

B. SCOPE OF THIS STUDY

The Statement of Work for this contract states its objective to be the "testing and evaluation of the Logoscan II Optical Reader (OCR) System to assess its potential for conversion of free-formatted multifont typeset Russian text to a computer processable format compatible with the Russian-English machine translation at Foreign Technology Division (FTD)."

Twenty pages of the book:

Доклады

Академии Наук СССР

were selected as representative of the general problem and are the subject of the present report. The book contains five separate fonts, viz., three title fonts, one main corpus font, and a bibliographical font. Due to the limitations of time and funding, Logoscan II has been optimized on the main corpus font only. Although all five fonts were placed in its memory and were used during the scanning operation, no effort has been made to increase the accuracy rate of these other four fonts. They can, of course, be brought to the same level of accuracy as the main corpus font at a later time. There is no inherent problem in the fonts themselves, nor in Logoscan II's ability to handle them. The primary aim has been to demonstrate a capability to scan such text with an accuracy rate and speed which would be significantly more economical than current manual techniques.

C. EQUIPMENT USED IN THIS STUDY

Many off-the-shelf OCR units were available; however, the ECRM 4500 proved to have the best combination of light source (a laser) and paper moving mechanism. A Data General S130 Eclipse was selected as the computer which processes the scanner's raw video data. The selection criteria in this case were: (1) compatibility with existing Logoscan I programs; and (2) a micro-code language which allowed Logos engineers to optimize certain high frequency program loops.

Logos engineers supervised the building and testing of a special purpose interface to control the functions of the 4500, which bypassed the 4500's internal PDP8 computer. This interface used a DMA channel for rapid transfer of information from the ECRM to the DG.

The functions of the ECRM are all controlled by the S130 using an interrupt driven operating system.

Appendix D contains a detailed list of the equipment used during this study.

D. TECHNICAL PROBLEMS

The four problem areas stated above are detailed here:

1. recognition of the character boundaries:

Since this book contains proportional spaced characters which are for the most part seraphed characters, a scanner operating at a 4 mil resolution frequently encounters touching characters. This would be true even at a 1 mil resolution, due to quality of certain sections of the typeset page (Page B1). The Logoscan II algorithm has been designed to search for clues as to the location of a boundary between touching characters even when they are proportionately spaced.

The upper and lower limits of a character, particularly when there are many ascenders and descenders (Page B4) are difficult to find. Line skewing compounds this problem. Once again, the algorithm because of its design can track a line and eliminate ascending or descending characters from contiguous lines.

It is this ability to define character boundaries in a proportionally spaced line, in the presence of noise, in a tight line spacing, and for in highly seraphed fonts that makes Logoscan II a fourth generation OCR device. The second characteristic which identifies the system as a new generation of OCR is its signature algorithm, viz., standard masking techniques are not used.

Characters such as the Russian H can be broken, that is, the two solid vertical strokes can appear to be two separate characters if the horizontal stroke is weak or missing. Logoscan II uses a hardware/software combination to solve this problem.

If the horizontal stroke is merely weak, good resolution will enable the system to see a stroke that might be missed with a less sensitive read head. To this end, Logos engineers modified the paper moving subsys em, i.e., the ECRM now has an effective resolution of 3 mils along the vertical axis of the page.

If the horizontal stroke is missing, however, Logoscan II relies on rules to examine areas around what looks like a piece of a character. If the system discovers a shape that, when added to the first piece, could be defined as a character, it will merge them and recognize a single character.

2. differentiating the character from all others in a reference alphabet:

The main corpus font contains unique characters, some of which look very similar to an OCR device, e.g., a Russian И and H. The method of signature analysis used by Logoscan II enables it to correctly differentiate such characters. The same character, such as a Russian И, when examined at different times during the scanning process, can vary (Page B7). The OCR must recognize every occurrence of this character correctly, even if it appears quite differently each time.

Characters can also vary from their normal appearance. This could be caused by pieces missing (usually seraphs) or noise that disguises the character's shape. The method of signature analysis used enables the system to differentiate between similar (but unique) characters, yet accurately identify characters, even if they vary in appearance.

Appendix C illustrates the system's processing of sub- and super-scripts.

3. line skewing and recognizing line boundaries:

Most OCR units introduce dynamic and static line skewing. This is caused by the paper transport system, or by the original positioning of text on the page. In addition, ascending and descending characters make the job of defining the boundaries of a line much more difficult (Page B4).

Logoscan II will read text that is dense, i.e., greater than 8 lines to the inch, with ascenders and descenders, and can track a line skewed by as much as one-half a line's height.

4. selecting the correct reference alphabet from a set of reference alphabets:

The five fonts noted above can appear on any page. The system tests a new line (if it was preceded by a blank line) against each font and selects the one having the lowest error count. Logoscan II can store in its main memory as many as ten such reference fonts. If more were needed, its secondary disk memory can be used and it could call in any number of secondary fonts.

Although the solutions to the above problems have been incorporated into Logoscan II software, there remains a set of problems related to the quality and content of the typeset page.

o Background Noise

As figure 4, Page B2, illustrates, the reverse side of a page shows through to the side being scanned. This background noise is general throughout the book used during this study. It cannot be eliminated, and its presence causes more than half of the errors experienced on any given page.

The thickness and quality of paper used in a book govern the degree of background noise. As the need and desirability to scan typeset material becomes more widely recognized, the problem of background noise could be largely eliminated by the proper selection of paper.

o Foreground Noise

Dirt, ink spots, and poor quality paper all introduce distortions to the character as seen by tne laser light source. Figure 3, Page B2.0, shows the sort of foreground noise experienced in this book.

o Identical Characters

A Russian H and an English H are the same character in the main corpus font. Only a post-processor program examining the character in context can differentiate such pairs. Page A4 lists the combinations appearing in this book.

E. GENERAL METHODOLOGY

After interfacing the ECRM/S130 system, Logos personnel concentrated on two main areas:

1. Recognition Speed

This activity required analysis of the Logoscan algorithms and supporting programs searching for high frequency loops that could be converted to S130 micro-code. Writing and debugging this code paralleled the second effort.

2. Recognition Accuracy

The five fonts were acquired and repeated testing of the main corpus font resulted in a fine tuned set of character signatures for this font. No such fine tuning was attempted on the other four fonts, nor were the Greek letters in the main corpus font optimized.

F. TECHNICAL RESULTS

The recognition speed on any given line is now 30 characters per second. The average recognition speed for a set of lines is 25 characters per second. This latter rate can be improved by foreground/background tasking of data acquisition and recognition at some later date.

The recognition accuracy for the main corpus font is now 98.6%. A high percentage of errors can be traced to the source document; although Logoscan II is designed to allow for wide variations within a character type, in this quality document, characters are sometimes so distorted as to be recognizable only in context.

G. IMPLICATIONS FOR FURTHER RESEARCH

The Logoscan II System is not now an operational system in a production environment. It could be made cost/effective for a number of Russian fonts with approximately three months of further effort, and it would be able to input any Russian font at a fraction of current keyboarding cost with nine months of further effort.

H. A PRODUCTION SYSTEM

Certain facts about Logoscan II in a production environment can be stated, and estimates of others are given below.

The system can read original pages or Xerox copies. Either way the preparatory work is about the same, viz.:

1. mark graphics as shown in Figure B5.

2. align one edge of paper at right angles to a line on the page. This would not be required of a book typeset in this country, but it is required for the book used in this test.

3. feed the pages 50 at a time into the scanner's paper feed tray.

The preparatory work for 50 pages is approximately one man-hour.

Reject Rate

Rejection rate corresponds to approximately .9 times the error rate. That is, 9 out of 10 errors can be flagged as errors. When Logoscan II is completed the estimated error rate for all fonts will be between 1% and 2% for Cyrillic fonts. This spread is a function of the quality of the original document. It would be 1% for the book used ın this study. Under these conditions there would be 22 flagged errors on a page, and given a good CRT editing system, they could be converted to correct characters in .041 man-hours.

Areas for Improvement

To achieve a full scale production system, certain new operational features would be required:

1. Additional special tests for problem characters.

2. A second ECRM unit interfaced to the S130. This would allow parallel processing of two separate books.

3. Post-editor programs to correct errors automatically by examining them in context.

4. A well thought out CRT editing system for Russian text.

As noted, about nine months of effort would achieve these goals.

Resultant Cost

Such a complete production system could be achieved in nine months for approximately $310,000,including all hardware, software and manpower costs.

Comparative Cost

Without current keyboarding cost accounting figures, no comparison can be made of automated vs. manual techniques. However, an estimate can be made of the cost for a 2500 character page using the full scale production system. The cost to achieve an edited (therefore nearly error-free) 2500 character page are:

1. System Costs:

Since the equipment will be user property after nine months, only monthly maintenance fees and power requirements are used as the basis for this estimate. In a one-shift, two-man operation approximately 800 such pages could be processed per day to the point where editors could begin correcting rejects. It would require 4.1 editing operators to correct rejects for these 800 pages. The system cost would be $.036 per page.

2. Manpower Costs:

As stated above, these 800 pages would require two operators and 4.1 editors per day, or .061 man-hours/page.

Appendix A

Доклады Академии наук СССР
1976. Том 230, № 6

УДК 547.963.3

ФИЗИЧЕСКАЯ ХИМИЯ

Член-корреспондент АН СССР Д. Г. КНОРРЕ,
С. Г. ПОПОВ, Т. А. ЧИМИТОВА

КИНЕТИЧЕСКИЕ ОСОБЕННОСТИ АФИННОЙ МОДИФИКАЦИИ БИОПОЛИМЕРОВ ДЛЯ РЕАКЦИЙ, ПРОТЕКАЮЩИХ С УЧАСТИЕМ АКТИВНЫХ ПРОМЕЖУТОЧНЫХ ЧАСТИЦ

Кинетические закономерности афинной модификации в настоящее время проанализированы для случая, когда реагент претерпевает превращение только находясь в комплексе с модифицируемым биполимером за счет существенного возрастания константы скорости взаимодействия реагирующей группы с модифицируемой группой биополимера в результате их пространственного сближения. В этом случае схема превращения записывается в виде

$$E+X \rightleftharpoons EX \rightarrow EZ,$$

где E – модифицируемый биополимер, X – афинный реагент, EX – комплекс биополимер – реагент, EZ – продукт модификации. При достаточном избытке реагента кинетика реакции описывается кинетическим уравнением для реакции первого порядка по биополимеру с кажущейся константой скорости, зависящей от концентрации реагента ([1], [2]). В работе ([3]) решение распространено на случай обратимой модификации, на случай, когда в процессе модификации биополимер проходит через несколько промежуточных состояний, и предложен приближенный метод решения для случая, когда концентрация биополимера величина того же порядка, что и концентрация реагента.

В настоящей работе рассматривается кинетика афинной модификации для случая, когда превращение реагента проходит через промежуточное образование активных частиц, причем образование этих частиц является лимитирующей стадией и в первом приближении не зависит от присутствия модифицируемого биополимера. С такими случаями приходится сталкиваться при фотоафинной модификации биополимеров, реагентами, содержащими азидную группу, когда лимитирующей стадией является превращение азидной группы в бирадикал нитрен ([4]) и при модификации производными ароматических 2-хлорэтиламинов, при которой лимитирующей стадией оказывается превращение 2-хлорэтиламиногруппы в реакционноспособный этиленимминиевый катион ([5]). В этом случае образовавшиеся активные промежуточные частицы могут либо модифицировать биополимер, либо реагировать с молекулами растворителя и другими низкомолекулярными компонентами раствора. В принципе не исключена возможность и неспецифической модификации биополимера. Для случая афинного (комплементарно-адресованного) алкилирования нуклеиновых кислот производными олигонуклеотидов, несущими остаток ароматического 2-хлорэтиламина, показано, что неспецифическая модификация вне комплекса нуклеиновая кислота – реагент происходит в незначительной степени ([6]). Для фотоафинных реагентов описаны случаи значительной неспецифичной модификации биополимера ([7]).

Doklady Akademii nauk SSSR
1976, Tem 230* # 6

UDK 547.96Z.Z

Clen.korreepeudeut AN SS6R D. G. KPORRE,
S. N PIOPOR, T. A. CIMI7*RA

KINETICESKIE OSOVENNOSTN AFINNO1 MODIFNKXQII (GO* S *CXSTIEM XKTPNN6P* PROMEJUTOCN6P*CXSTIQ

Kineticeskiv zakonomernosti afinno1 modifikaqii v nasto45ee vrem4 proanalizirovany dl4 sluca4, ko5a reagent preterpevaet prevra5enie tol6ko naxod4s6 v komplekse s modifiqiruem6*1 bipolimerom za scet su5estvennogo vozrastani4 konstanty skorosti vzaimode1stvi4 reagiruh5el gruppy s modifiqiruemo1 gruppo1 biopolimera v rezul6tate ix pro stranstvennoro sblijeni4. V 3tom slucae sxema prevra5eni4 zapisyvaets4 v vide

E+X=EX>EZ.

gde E modifiqiruemy1 biopolimer, X- afinny1 reagent, EX- kompleks biopolimer-reagent, EZ -produkt modifikaqii. Pri dostatocnom izbytke reagenta kinetika reakqii opisyvaets4 kineticeskim uravn*niem dl4 reakqii pervogo por4dka po biopolimeru s kaju5e1s4 konstanto1 skorosti, zavis45e1 ot konqentraqii reagenta (1, 2). V rabote (3) rewenie rasprostraneno na sluca1 obratimo1 modifikaqii, na sluca1, kogda v proqesse modifikaqii biopolimer proxodit cerez neskol6ko promejutocnyx sosto4ni1, i predlojen priblijenny1 metod reweni4 dl4 sluca4, kogda konqentraqi4 biopolimera velicina togo je por4dka, cto i konqentraqi4 reagenta.

V nasto45e1 rabote rassmatrivaets4 kinetika afinno1 modifikaqii r14 sluca4, kogda prevra5enie reagenta proxodit cerez promejutocnoe obrazovanie aktivnyx castiq, pricem obra3ovanie 3tix castiq 4vl4ets4 limitiruh5e1 stadie1 i v pervom priblijenii ne zavisit ot prisutstvi4 modifiqiruemogo biopolimera.Stakpmisluca4miprixodits4stalkivat6s4 pri fotoafinno1 modifikaqii biopolimerov. reagentami, soderja5imi azidnuh gruppu, kogda limitiruh5e1 stadie1 4vllets4 prevra5exie a3idno1 gruppy v biradikal nitren (4) i pri modifikaqii proizvodnymi aromaticeskix 2-xlor3tilaminov, pri kotoro1 limitiruh5e1 sta die1 okazyvaets4 prevra5enie 2xlor3tilaminogruppy v reakqi*nnosposobny1 3tilenimmonievy1 kation (3). V 3tom sxucae obrazovavwies4 aktivnye promejutocnye castiqymogutlibomodifiqirovat6biopolimer, libo reagirovat6 s molekulami rastvoritel4 i drugimi ni3komolekul4rnymi komnonentami rastvora. V prinqipe ne isklhcena vozmojnost6 i nespeqificesko1 modifikaqii biopolimera. D14 s1Gca4 afinnogo (komplementarno-adresovannogo) alkilirovzni4 nukleinovyx kislot proizvodnymi oligonukleogidov, nesu5imi ostatok aromaticeskogo 2xlor3tilamina, pokazano, cto nespeqificeska4 modifikaqi4 vne kompleksa nukleinova4 kislota -reagent proisxodit v nezxacitel6no1 stepeni (6). D14 fotoafinnyx reagentov opisany s4ucai znacitel6no1 4esneqificno1 modifikaqii biopolimera (7).

RUSSIAN-ENGLISH ALPHABET EQUIVALENTS

Lower Case

Russian	English	Russian	English
а	a	р	r
б	b	с	s
в	v	т	t
г	g	у	u
д	d	ф	f
е	e	х	x
ж	j	ц	q
з	z	ч	c
и	i	ш	w
й	1	щ	5
к	k	ъ	7
л	l	ь	6
м	m	ы	y
н	n	э	3
о	o	ю	h
п	p	я	4

Upper Case

Russian	English	Russian	English
А	A	Р	R
Б	B	С	S
В	V	Т	T
Г	G	У	U
Д	D	Ф	F
Е	E	Х	X
Ж	J	Ц	Q
З	Z	Ч	C
И	I	Ш	W
Й	1	Щ	5
К	K	ъ	7
Л	L	Ь	6
М	M	Ы	Y
Н	N	Э	3
О	O	Ю	H
П	P	Я	4

SPECIALS = @	=	>	<	#
ψ α η κ δ Å ξ β σ φ ζ ε π τ √ ∞ ∞ (SUB-SCRIPT)	± ≈ ≅ ≃ ⇄	⩾ > →	⩽ <	3/2 №

Characters with similar shapes in the Cyrillic and Roman alphabets

Lower Case

Roman	Cyrillic
y	y (u)
e	e (e)
r	r (g)
a	a (a)
p	p (r)
o	o (o)
c	c (s)
m	m (m)
x	x (x)

Upper Case

Roman	Cyrillic
A	A (A)
B	B (V)
C	C (S)
E	E (E)
H	H (N)
K	K (K)
M	M (M)
O	O (O)
P	P (R)
T	T (T)
Y	Y (U)
X	X (X)

Appendix B

A close study of the Russian word in Fig. 1 provides a good example of the difficulties in scanning this type of document and the methods employed by Logoscan II to overcome these difficulties.

аддукты

Figure 1

Figure 2(A) is a "normal" character as seen by the scanner, i.e., positioned correctly in relation to the other characters, and not touching surrounding characters. The next three characters, "ДДУ", all contain descenders and all bleed into each other, Figure 2(B). The system first "pushes" the characters up, Figure 2(C), then makes a first effort at separation, Figure 2(D). The system then decides that this shape, Figure 2(D), is still two characters and separates them further, this time recognizing the character as "Д", Figure 2(E).

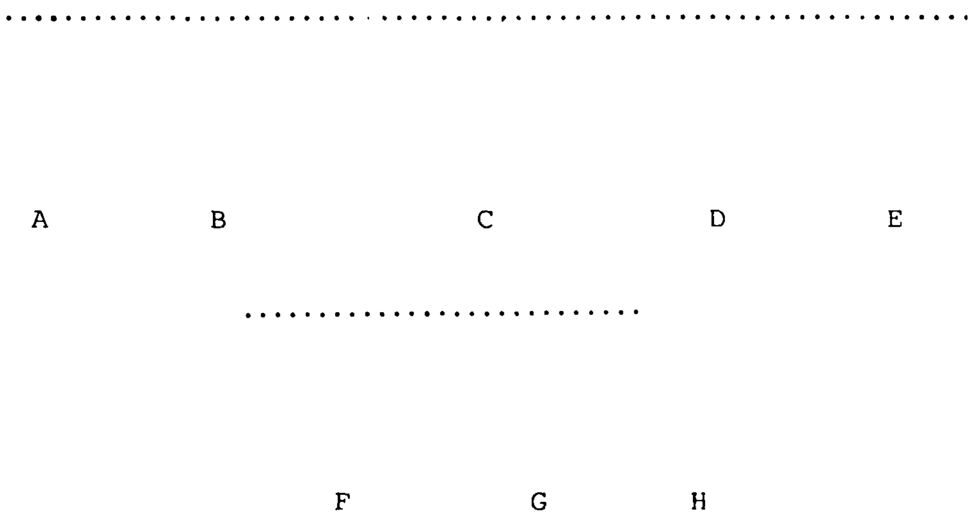
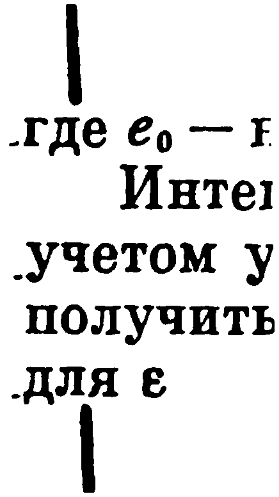


Figure 2

The procedure is repeated to define the next "Д", Figure 2(F), (G), (H). Note that processing is restarted from the original positioning of the characters. This is so that we can accurately track a line of text, even with skewing.

When scanning typeset material, the OCR will be presented with a variety of noise present on the page. This can be caused by ink splattered on the platen (Fig. 3 to left of characters) or impressions from the reverse side of the page (Fig. 4), or chips and pieces of type suspended between characters (Fig. 5).



где e_0 — г

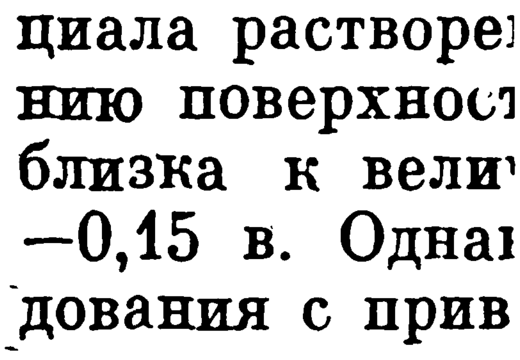
Инте

учетом у

получить

для ε

Figure 3



циала растворе

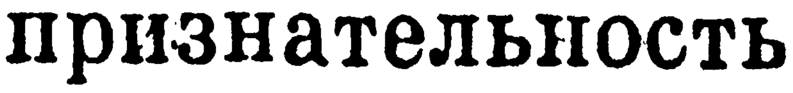
нию поверхнос

близка к вели

—0,15 в. Одна

дования с прив

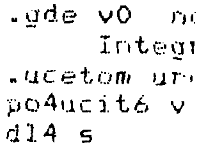
Figure 4



признательность

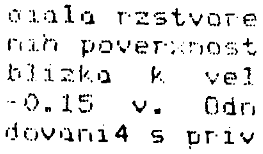
Figure 5

For the above cases Logoscan II was able to correctly handle the difficulties (Figs. 6, 7, 8). However, in certain cases the noise was too severe, and the character was distorted (Fig. 9). Characters such as this represent 50% of the error rate.



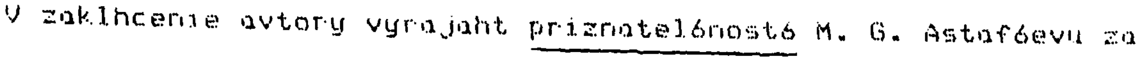
```
.gde v0  nc
     Integı
.ucetom ur
po4ucit6 v
d14 s
```

Figure 6



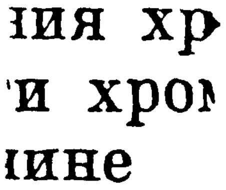
```
oıala rzstvore
nah poverxnost
blizka  k  vel
-0.15  v.  Odn
dovani4 s priv
```

Figure 7



V zaklhcenıe avtory vyrajaht priznatel6nost6 M. G. Astaf6evu za

Figure 8



ıия хр
'и хром
ıине

Figure 9

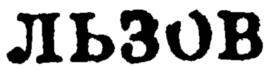
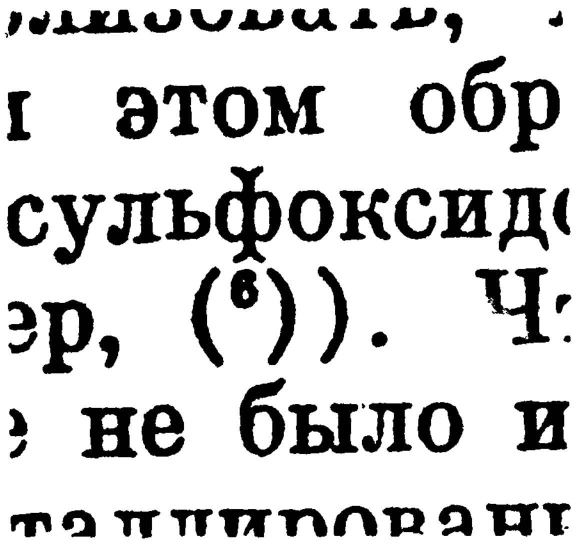


Figure 10

Characters formed by hot type can be deformed due to missing pieces. While this usually occurs as weak strokes, sometimes a chipped slug, or ink missing from a part of the slug, produces characters such as in Fig. 10.

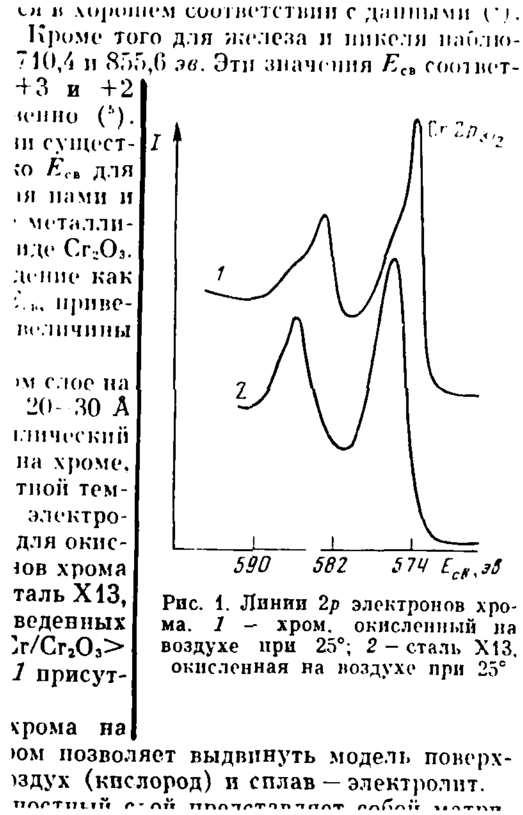
Ascending and descending characters are major problems for OCR devices, particularly at 7.7 lines per inch. Difficult situations such as the one pictures in Fig. 11 can occur. The sequence, left parenths, super 6, right parenths, right parenths, is dangerously close to both the bottom portion of the Cyrillic "ф" on the line above, and the top portion of the Cyrillic "б" on the line below. Logoscan II was able to handle areas like this because it can "track" a line of text and make decisions on what areas to examine.



I этом обр
сульфоксид(
эр, (^{6})). Ч
; не было и

Figure 11

To handle graphics Logoscan II requires that the operator draw a line to the left of the graphics area (Fig. 12). The system will then recognize this line as a delimiter and go to the next line of text. In the output, the system leaves proportional white space for insertion of graphics later (Fig. 13).



ся в хорошем соответствии с данными ( ).
Кроме того для железа и никеля наблю-
710,4 и 855,6 *эв*. Эти значения $E_{св}$ соответ-
+3 и +2
ченно ([5]).
ии сущест-
ко $E_{св}$ для
ия нами и
металли-
иде Cr_2O_3.
дение как
$E_{св}$, приве-
величины

ом слое на
20- 30 Å
ллический
на хроме,
тной тем-
электро-
для окис-
ионов хрома
таль X13,
веденных
$Cr/Cr_2O_3>$
1 присут-

хрома на
ом позволяет выдвинуть модель поверх-
оздух (кислород) и сплав — электролит.

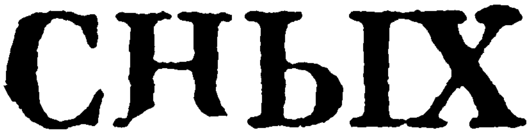
Рис. 1. Линии 2*p* электронов хрома. *1* — хром, окисленный на воздухе при 25°; *2* — сталь X13, окисленная на воздухе при 25°

Figure 12

```
d4ts4 v xorowem sootvetstvii sdannymi (4).
 Krome togo dl4 jele3a i nikel4 na6lh-
710,4 i 855,6 3v. 3ti 3naceni4 E6v sootvet-
   +3  i  +2
venno  (5).
 oni su5est-
ko Esp dl4
na4 nami i
t ne metalli-
v vide Sg203.
adenie kak
  Esv, prive-
ol veliciny
(8).
nom sloe na
20-30 @
etalliceskil
cto na xrome,
natnol tem-
# 3lektro-
ki dl4 okis
onov xroma
-stal6 X13,
edennyx
ie Sg)Sg203>
 1 prisut-

roma  na
m pozvol4et vyqvinut6 model6 poverx-
ozdux (kislorod) i splav -3lektrolit.
```

Figure 13



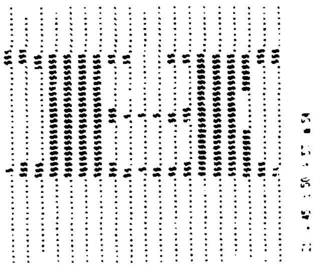
СНЫХ

Figure 14

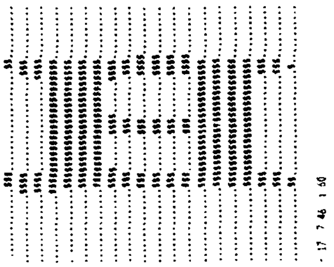
The sequence "Ы*" (-Y*- where * represents any character) posed a special problem for Logoscan II if the character following the "Ы" bleeds into it (Fig. 14). Since Logoscan II looks for characters when presented with a group of touching characters, and the shape "ь" is a legitimate Cyrillic character, the I portion might be taken as part of the next character.

This situation can only be corrected by post-processing using grammatical rules, which Logos plans to implement at a later date.

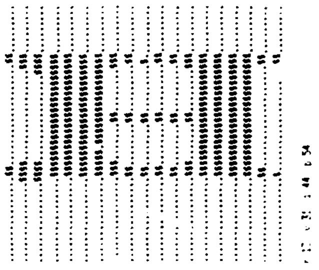
The printout below demonstrates how much a typeset character can vary. Each of the figures (A, B, C) is an image of what the read head "sees." The system has to identify each of these as the same character, despite the varying line thicknesses.



A



B



C

Appendix C

Logoscan II is able to handle subscripts and superscripts as shown below:

INPUT:

щимися соединениями (^{13}). Если полученные нами экспериментальные

OUTPUT:

```
5imis4 soedineni4mi (13). Esli polHevnye nami 3kspeгimevtal6vye
```

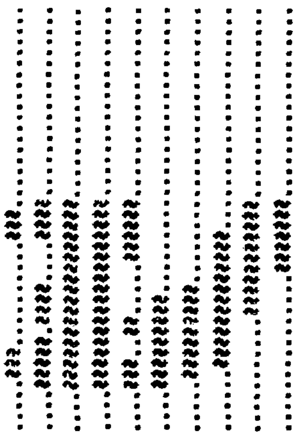
INPUT:

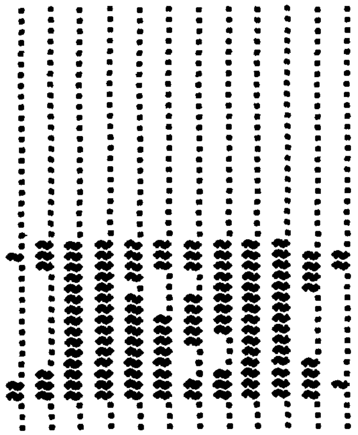
турой типа флюорита общей формулы $Nd_{1-x}Te_xO_{1+x}F_{1-x}$. Они сущест

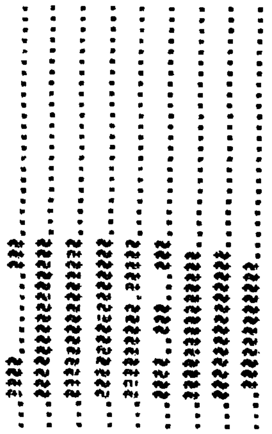
OUTPUT:

```
turol tipa flhorita ob5el formuly Nd1-xTexO1+xF1̄-x. Oni su5est
```

Superscripts and subscripts are somewhat more difficult than normal characters in the main corpus font because of their lack of definition. The three pictures below are subscripts as seen by the scanner at 3 by 4 mils resolution.







Appendix D

DATA GENERAL

ITEM	QTY	DESCRIPTION
		HARDWARE:
		CPU and Related Items
DG/8611-I	1	Eclipse S/130 computer with 64KB MOS memory, battery backup and ERCC
DG/8615	1	Writable control store
DG/8614	1	Character Instruction Set
DG/1012P	1	One-bay cabinet (240V @ 50A)
		Console, Editing Terminals and Related Items
DG/4075	2	I/O Interface Subassembly
DG/4078	2	EIA (RS232C) Interface
DG/6052-D	1	CRT Display Terminal (24X80)
DG/6086	1	180 cps, bidirectional 9X7 dot matrix printer
DG/4079	1	Real Time Clock
		Disc Subsystem and Related Items
DG/6070	1	20 Mbyte DGC Cartridge Disc Subsystem, (10 fixed, 10 removable) includes cartridge.
		Interface
DG/ECRM4500	1	Direct memory access: DG/8611-I and ECRM/4500 Control: DG/8611-I and ECRM/4500
		SOFTWARE:
		Eclipse Stand-Alone Operating System, Assembler, Macro-Assembler, Fortrain IV, Basic Algol Compilers and System Utilities
		Eclipse RDOS Operating System Assembler, Macro Assembler Fortran IV, Basic Algol Compilers, Sort/Merge, CSP & System Utilities
		Eclipse RTOS Real Time Operating System
		Eclipse Operating System for Eclipse Series
		Diagnostic Operating System for Peripherals

ECRM

ITEM	QTY	DESCRIPTION
		HARDWARE:
ECRM/4500	1	4000 Series Autoreader